

EXHIBIT A

11/13/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential

Michael Patrick Clark 30(b)(6), Vol I & Vol II

Page 1

IN THE UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

RICHARD KADREY, et al.,)	
)	
Individual and Representative)	Lead Case No.
Plaintiffs,)	
)	3:23-cv-03417-VC
v.)	
)	
Meta Platforms, Inc.,)	
)	
Defendant.)	
)	

** H I G H L Y C O N F I D E N T I A L **

30(b)(6) VIDEOTAPED DEPOSITION OF
META PLATFORMS, INC.

BY: MICHAEL PATRICK CLARK

Denver, Colorado

VOLUMES I AND II

Wednesday, November 13, 2024

Thursday, November 14, 2024

Reported stenographically by:

Michelle Kirkpatrick, RDR-CRR-CRC-CRI, FCRR

DIGITAL EVIDENCE GROUP
1730 M Street, NW, Suite 812
Washington, D.C. 20036
(202) 232-0646

Page 53

1 we were doing on privacy to make sure that the model
2 would not regurgitate information about private
3 individuals.

4 We believed that, and it is our position
5 that using copyrighted data for training falls
6 within fair use boundaries; but while we had the
7 mitigations that we had already built for privacy,
8 we applied and made sure that we were doing the same
9 mitigations more broadly, including, in both privacy
10 and IP, measuring the likelihood of memorization or
11 regurgitation or re-identifiability of content.

12 Q Okay. And tying that back to your
13 testimony about transformation, is it Meta's
14 position that transformation occurs when the data
15 that is ingested is tokenized?

16 A It is the combination of the entire
17 process.

18 So it is -- are you --

19 Q No, I'm waiting for your --

20 A I thought you were asking something.

21 It's the entire process. It's -- there's
22 an entire set of data, and then that data is cleaned

1 up. It has duplicates removed. It has unneeded
2 characters, repeating characters. It has metadata.
3 It has other things removed to get down to a base
4 clean, parsed dataset.

5 That parsed dataset then goes through a
6 series of evaluations to find out which combination
7 and what token size, what training passes, which
8 methods, and research that we do at that stage to
9 identify what is best.

10 Once we get through that process, we then
11 go into the full training process. So we take what
12 were the best results and move into the tokenization
13 process and the weights process and combinations of
14 how the final model, from a pre-trained perspective,
15 works and functions.

16 We then go through an alignment phase to
17 train the model for specific kinds of capabilities
18 based on pre-training and post-training, and then
19 whatever additional fine-tuning is needed for the
20 specific purpose of the model.

21 And the entire life cycle of that is what
22 we measure for memorization and believe that,

Page 55

1 through that entire process, it is transformative.

2 Q Okay. Is it Meta's position that scraping
3 data from websites and making copies of that data is
4 transformative?

5 MS. HARTNETT: Objection to the form and
6 to the extent it calls for a legal conclusion.

7 A Can you clarify what you mean by
8 "scraping"? That's a broad --

9 BY MS. POUEYMIROU:

10 Q So scraping might not be the right term
11 here. But you described this pipeline, and it
12 seemed to start from the processing of data that's
13 already been copied and create -- a dataset has been
14 made, and now you are preparing it for ingestion.

15 My question is -- and you described -- is
16 it correct to say that you're saying that the
17 preparation of the data that's been copied begins
18 the process of transformation?

19 MS. HARTNETT: Objection to the form and
20 to the extent it calls for a legal conclusion.

21 MS. POUEYMIROU: I'm -- I am merely using
22 the language that Mr. Clark is using.